# Извештај за практично истраживање

|  |
| --- |
| **Истраживачи: име, презиме, број индекса (највише четири особе, ред по особи)** |
| Тамара Ранисовић, IN 1/2020 |

|  |
| --- |
| **Шифра истраживања: кодни назив (једна или две речи)** |
| Pantheon Project |

|  |
| --- |
| **Наслов истраживања: сажето исказана суштина истраживања (до 15 речи)** |
| Експлоративна анализа података о историјским личностима |

|  |
| --- |
| **Предмет, циљ, методи и очекивани резултати истраживања (бар 100 речи)** |
| Предмет овог истраживања је експлоративна анализа над скупом података о историјским  личностима. Овај скуп података настао је у оквиру пројекта Пантеон. Прукпљени су  подаци о историјским личностима (место и година рођења/смрти, занимање, број језика  чланака, укупан број прегледа свих Википедија чланака, просечан број прегледа чланака  по Википедија издању...) који имају чланке у више од 15 језичких издања Википедије. На  крају, за сваку историјску личност израчунат је индекс популарности (енг. Historical  Popularity Index - HPI), на основу 5 различитих параметара: „старост“ историјске личности  (од рођења до сада), број језичких издања Википедије у којој личност има присуство, број  прегледа чланака на различитим језицима, стабилност броја прегледа чланака током  времена, и број прегледа чланака на свим неенглеским језицима. Мотивација за одабир  ове теме истраживања лежи у жељи за проналаском одговора на следећа питања: Које  историјске личности су оствариле највећи HPI? Када и где су живеле и чиме су се бавиле?  Како се разликују занимања важних историјских личности из, нпр. средњег века и 21.  века? Како се креће број прегледа чланака по одређеним занимањима, државама или језику  чланка?  За анализу података користиће се програмски језик Pythоn. Биће извршено чишћење  података, преглед дескриптивне статистике одређених обележја, анализе обележја уз  груписање по полу историјских личности, корелациона анализа између одређених  обележја, визуализација значајних резултата...  Очекује се да ће највећи HPI остварити историјске личности које су имале значајне  доприносе у виду научних и техничких открића, док се за 21. век очекује да ће то бити  личности које већ имају велики број пратилаца на друштвенимм мрежама. Такође, очекује  се да ће током историје бити приметна промена у занимањима са високом вредношћу за  HPI. Поред тога, очекује се да је велик број историјских личности из старог века живело у  области данашње Грчке и Италије, док они млађе „старости“ долазе из различитих делова  света. |

|  |
| --- |
| **Опис коришћених скупова података (бар 30 речи по скупу података)** |
| https://www.kaggle.com/datasets/mit/pantheon-project  Подаци су прикупљени у оквиру пројекта Пантеон, који је започела MIT Collective  Learning група, док је за развој данас одговорна компанија Datawheel. Идеја овог пројекта  је коришћење биографских података како би се уочили шаблони људске колективне  меморије. Скуп података има 17 колона, међу којима су: пуно име историјске личности,  „старост“, пол, година, град, област, држава и континент рођења, географска ширина и  дужина места рођења, занимање, индустрија и област индустрије, број различитих језика  чланака, укупан број прегледа чланака, просечан број прегледа по Википедија издању, и  последња колона која представља HPI вредност. Скуп садржи податке за 11328  историјских личности. |

|  |
| --- |
| **Опис обављених активности и њихове расподеле међу истраживачима (бар 150 речи)** |
| Обављено је:   * разумевање података * чишћење података * попуњавање одређених недостајућих вредности (нпр. континент) * приказ историјских личности сортираних по:   + највећем броју прегледа чланака   + највећем просечном броју прегледа чланака   + број различитих језика чланака   + HPI * приказ претходно сортираних личности према полу * приказ значајних личности из Србије, такође сортирани по истим параметрима * приказ расподеле по полу, индустрији, области индустрије, континентима, државама, градовима, историјском добу * приказ односа особа женског и мушког пола по области индустрије * приказ индустрија и области индустрија за Топ 100 по HPI вредности * приказ HPI у односу на годину рођења и област индустрије * приказ просечног броја прегледа чланака у односу на годину рођења и област индустрије * приказ промена области индустрије током различитих историјских доба * приказ промена континента порекла личности током различитих историјских доба * приказ расподеле HPI вредности по областима индустрије * приказ броја личности у Топ 15% са највећим HPI по вековима |

|  |
| --- |
| **Кратак опис претходних истраживања која су изведена над коришћеним скуповима података од стране других особа и приказ главних разлика између садашњег и претходних истраживања (бар 70 речи)** |
| Претходна истраживања су у великој мери покрила највећи део анализе овог скупа података. Анализе које су претходни истраживачи урадили покушала сам и ја да рекреирам, користећи друге библиотеке, и добила исте резултате. Међутим, постоје и одређени прикази које нисам видела у радовима које сам истраживала: приказ расподеле по континентима, приказ значајних историјских личности из Србије, посебан приказ значајних обележја за особе женског и мушког пола, приказ процената личности које припадају сваком од историјских доба, приказ расподеле области индустрија током историјских доба... |

|  |
| --- |
| **Анализа испуњења циља и остварења очекиваних резултата (бар 30 речи)** |
| Потврђене су хипотезе:   * Историјске личности са највећим имају HPI значајне доприносе човечанствз, али не искључиво у виду научних и техничких открића – присуне су личности из области филозофије, религије... * Приметна је промена у занимањима током различитих историјских доба * Приметна је значајна разлика у погледу личности које имају највећи број прегледа чланака у односу на оне које имају висок HPI. |

|  |
| --- |
| **Могућности за примену истраживања у пракси (бар 30 речи)** |
| Методе и технике које сам применила при овом истраживању могу се користити у било којој даљој експлоративној анализи података. У пракси се конкретно може применити логика која стоји иза одређених пита дијаграма, расподела, приказа сложенијих графика итд. Такође, истраживање се показало корисним јер се добија шира слика о развоју историје које ће значити уколико се неко одлучи да примени методе науке о подацима у друштвеним наукама или слично. |

|  |
| --- |
| **Идеје за побољшање истраживања (бар 30 речи)** |
| Истраживање се може побољшати тако што ће се одрадити детаљније анализе обележја као што су географска ширина и дужина и размотрити да ли постоји одређена веза између њих и HPI вредности. Такође, може се одрадити додатна анализа личности из касног 20. и 21. века, установити ко су одређене личности, чиме се баве, да ли је популарност ових личност повезана са њиховим утицајем које имају на друштвеним мрежама... Истраживање је могуће обогатити комбинацијом са другим скуповима података који су из сличне области попут историје. |